

2015

Trabajo Final

Postgrado – Facultad de Informática - UNLP

Curso: Visualización

Carrera: Especialización en Computación Gráfica, Imágenes y
Visión por Computadora

Profesores: Dra. Silvia Castro, Dr. Martin Larrea

Alumno: Lic. Carlos Sebastián Castañeda

Tema: Implementación de una herramienta para visualización
de Coordenadas Paralelas con HTML 5, Javascript y D3



Tabla de contenido

| | |
|---|----|
| Tabla de contenido | 2 |
| Introducción..... | 3 |
| Coordenadas Paralelas..... | 3 |
| Brushing | 3 |
| Limitaciones | 4 |
| D3..... | 4 |
| Visualizador de Coordenadas Paralelas | 5 |
| Implementación | 5 |
| Utilización | 6 |
| Formato de los datos | 7 |
| Instalación..... | 7 |
| Trabajos a futuro..... | 8 |
| Soporte para datos ordinales..... | 8 |
| Soporte para nuevos formatos de entrada | 9 |
| Mejorar el manejo de errores | 9 |
| Exportar la visualización en diferentes formatos..... | 9 |
| Mejorar el styling general de la visualización..... | 9 |
| Conclusiones | 9 |
| Referencias | 10 |

Introducción

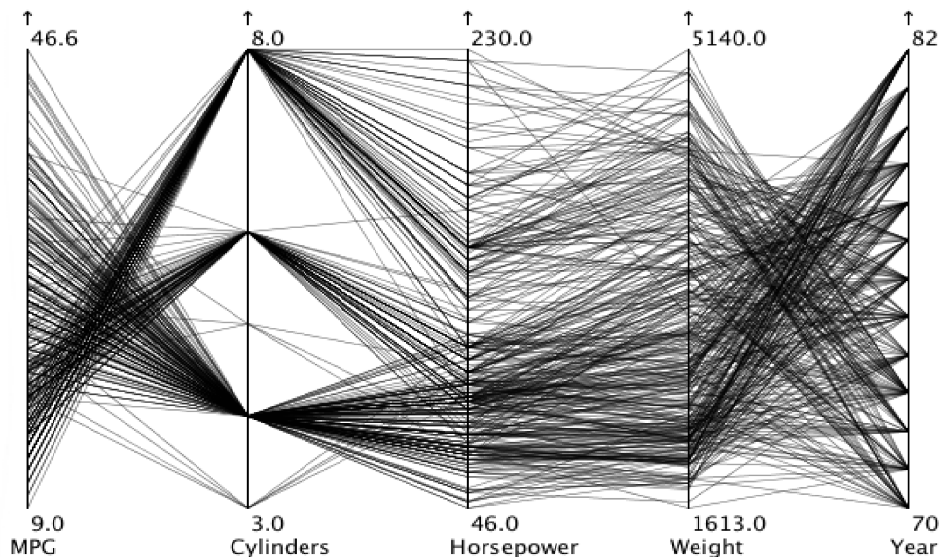
Existen numerosos ejemplos de visualizaciones realizadas con la técnica de Coordenadas Paralelas [1] en la Web. Sin embargo, todos los ejemplos muestran un juego de datos fijos de un dominio particular.

Por recomendación de la cátedra de “Visualización”, surgió la idea de implementar un visualizador/herramienta de Coordenadas Paralelas que permita cierto dinamismo. Es decir, poder ingresar los datos a visualizar, obteniendo así una herramienta que rápidamente permita visualizar un conjunto de datos con esta técnica; lo cual podría derivar en alguna conclusión sobre los mismos o determinar que la técnica de Coordenadas paralelas no es la más adecuada para dichos datos.

Coordenadas Paralelas

Las coordenadas paralelas son una de las técnicas de visualización más famosas, y mas nombrada en *papers* relacionados con visualización [1].

Las coordenadas paralelas son una alternativa para la visualización de geometría n-dimensional en el plan, que permite analizar datos multivariados. Un conjunto de puntos en un espacio n-dimensional se muestra en un sistema de coordenadas representado mediante n líneas paralelas, típicamente verticales y equi-espaciadas. Un punto en ese espacio n-dimensional se representa como una poli-línea con vértices sobre los ejes paralelos [2]. La imagen 1, muestra un ejemplo de datos visualizados con coordenadas paralelas.

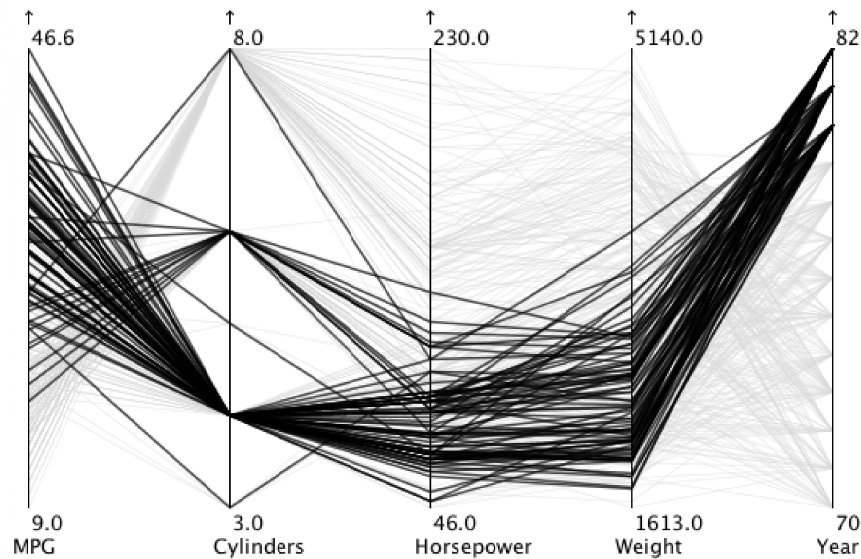


1: visualización utilizando la técnica de coordenadas paralelas [1]

Brushing

Además de necesitar un poco de experiencia en la lectura de una visualización de coordenadas paralelas, la mejor manera de adquirir conocimiento de un juego de datos, que utiliza esta técnica, es claramente la interacción. La principal interacción en una visualización de

coordenadas paralelas es la técnica de “brushing”, por razones que son obvias al observar la imagen 2, para que dicha interacción tenga sentido, debemos observar todos los ejes en conjunto.



2: Interacción de “Brushing” sobre una visualización de coordenadas paralelas [1]

Limitaciones

Por supuesto que una visualización de coordenadas paralelas tiene sus limitaciones. Cuando el número de elementos de datos es muy alto, se produce un efecto de *overplotting*¹ que puede dificultar la visualización. Mientras que el número de dimensiones en pantalla también debe ser por debajo de aproximadamente 12, para que la visualización como un todo tenga sentido, cualquiera de estos dos parámetros por encima de dichas cotas será difícil de leer.

Existe mucho trabajo en visualización sobre reordenamiento automático de los ejes, *clustering* de ejes similares, etc. Que facilitan la visualización de datos con un alto número de dimensiones. Pero en general, la técnica de coordenadas paralelas funciona mejor para juegos de datos con un moderado número de dimensiones y no más de unos pocos miles de registros.

Por último cabe destacar que la técnica no funciona muy bien para datos categóricos o con pocos valores por eje.

D3

D3 (Data-Driven Documents) es una librería Javascript para la manipulación de documentos basados en datos. D3 nos permite “dar vida” a un determinado conjunto de datos utilizando HTML, SVG y CSS. La librería pone énfasis en la utilización de los estándares Web haciendo uso de todas las capacidades de un navegador Web moderno, sin atarse a ningún *framework* propietario, combinado potentes componentes de visualización y un enfoque dirigido por los datos para la manipulación del DOM (*Document Object Model*) [3].

¹ En algunos gráficos, especialmente aquellos que utilizan puntos o líneas para codificar datos, múltiples objetos pueden terminar compartiendo el mismo espacio, posicionándose uno arriba de otro. Esto dificulta o hace imposible ver los valores individuales, que pueden terminar perjudicando el análisis de los datos. Este problema es conocido como “overplotting” [16].

D3 permite enlazar datos arbitrarios al DOM, y luego aplicar transformaciones dirigidas por datos al documento. Por ejemplo, podemos utilizar D3 para generar una tabla HTML desde un arreglo de números. O, utilizar los mismos datos para crear un grafico de barras SVG interactivo con transiciones e interacciones.

D3 no es un *framework* monolítico que intente proveer cada característica concebible. En cambio, resuelve la raíz del problema: *manipulación eficiente de documentos basados en datos*. Esto evita representaciones propietarias y ofrece una extraordinaria flexibilidad, exponiendo todas las capacidades de los estándares Web tales como HTML, SVG y CSS. Con un mínimo de *overhead*, D3 es extremadamente rápido, soporta grandes volúmenes de datos y comportamientos dinámicos para interacciones y animaciones. El estilo funcional con el que está desarrollado permite la reutilización de código a través de una colección de diversos componentes y *plugins*.

Visualizador de Coordenadas Paralelas

El principal objetivo de la herramienta implementada fue la de poder visualizar rápidamente un conjunto de datos con la técnica de coordenadas paralelas. Existen varios ejemplos de gráficos de coordenadas paralelas en la Web implementados con D3, pero todos ellos limitados a un conjunto de datos particular. A su vez, como objetivo secundario se busco que la utilización de la misma sea simple; y que los requerimientos para su instalación sean mínimos.

Implementación

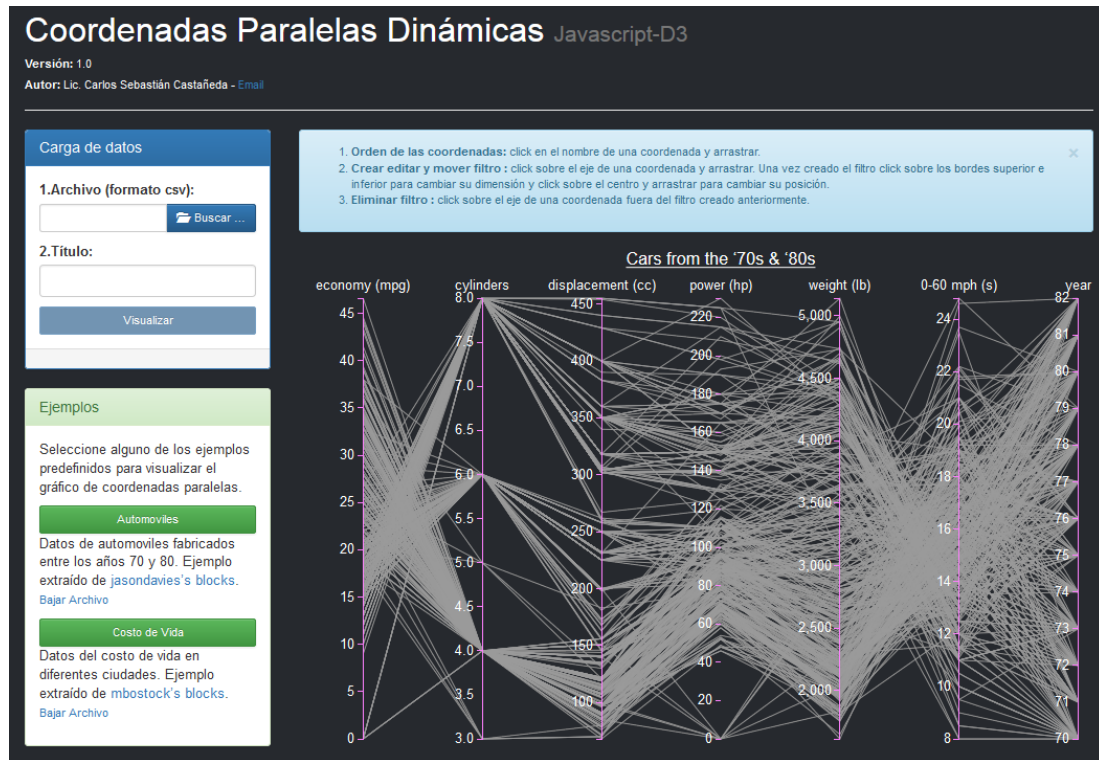
Por los motivos antes mencionados, la implementación de la herramienta se realizo utilizando únicamente D3, HTML5, CSS y Javascript. Sin la necesidad de contar con un servidor Web con demasiados requerimientos, más que permitir atender requerimientos HTTP simples. La herramienta podría utilizarse íntegramente en forma local a excepción de los ejemplos que la misma brinda los cuales se completan mediante un requerimiento HTTP, debido a restricciones de algunos navegadores modernos como *Google Chrome* que no permiten hacer *request* a dominios locales [4].

Para la implementación de la herramienta se utilizaron las siguientes librerías y recursos:

- *D3*: para la implementación del grafico de coordenadas paralelas y acceso a los elementos DOM.
- *HTML 5*, *Bootstrap* [5] [6], *JQuery*: para el *styling* general de la página WEB e interacciones adicionales.
- *Python SimpleHTTPServer* [7]: para probar la funcionalidad de los ejemplos predefinidos que requieren de un servidor HTTP para generar los *requests*.
- La idea, código y *styling* general de la herramienta se baso en varios ejemplos Web; refactorizando luego dicho código según las necesidades específicas [8].
- El procesamiento de los archivos de datos se realiza íntegramente con Javascript en el lado del cliente [9]. Por este motivo es que la herramienta puede utilizarse en forma local.

Utilización

La imagen 3 muestra la interfaz de la herramienta para generación de gráficos de coordenadas paralelas **DPC (Dynamic Parallel Coordinates)**. Las siguientes secciones describen la forma de cargar y visualizar datos con DPC.



3: Interfaz de DCP Herramienta para la generación de graficos de coordenadas paralelas

Visualizar un archivo

Para visualizar un archivo seleccionar de la sección “Carga de Datos” el archivo en formato CSV deseado, ingresar un titulo para el grafico [10] (opcional, en cualquier caso se puede dejarlo vacio) y clickear “Visualizar” para construir y mostrar la visualización.

Ver un ejemplo predefinido

La herramienta incluye dos ejemplos predefinidos [8] [11]. Para visualizarlos seleccionar el ejemplo deseado en la sección “Ejemplos” y clickear el botón correspondiente a cada ejemplo. Por defecto al cargar la herramienta se muestra el primero de estos ejemplos.

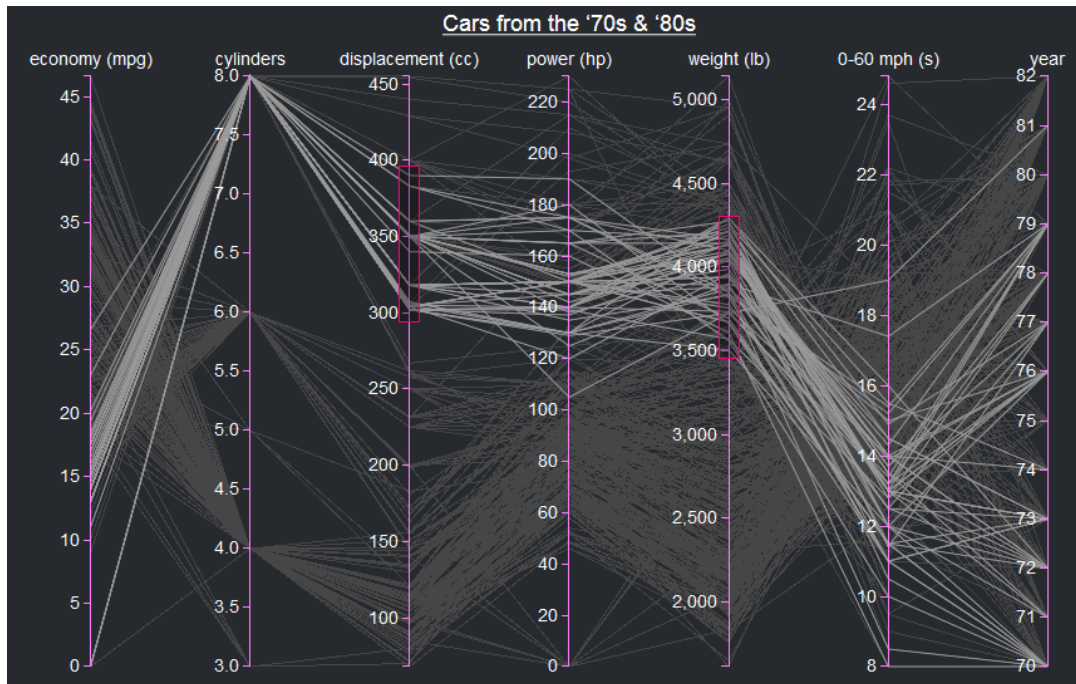
Interacción con la visualización

El grafico de coordenadas paralelas mostrado implementa varias interacciones:

- **Brushing:** es posible filtrar los datos del gráfico de coordenadas paralelas mediante la interacción de brushing. Para ello, se debe clickear y arrastrar sobre una dimensión para

crear un filtro. Dado un filtro ya creado es posible cambiar su tamaño o su posición o eliminarlo haciendo click fuera del mismo sobre el eje de la dimensión donde se creó. La imagen 4 muestra un ejemplo de interacción mediante “brushing”.

- **Reordenar los ejes:** es posible reordenar las dimensiones o ejes del gráfico haciendo *click* sobre las etiquetas de cada eje y arrastrando para obtener una nueva visualización de los datos.



4: DPC interacción “brushing”, filtrado de los datos en una o más dimensiones

Formato de los datos

La herramienta permite cargar y visualizar archivos en formato CSV [12], con los siguientes requisitos:

1. La primera línea del archivo es utilizada para el nombre de las etiquetas de cada eje.
2. La primera columna del archivo representa cada elemento de dato individual. En otras palabras, se traduce en una poli-línea del gráfico.
3. Un archivo con N columnas se visualizará con N-1 dimensiones (ejes).
4. La primera columna será en general de tipo String y el resto deben ser numéricas.

La imagen 5 muestra un ejemplo de archivo CSV válido para la herramienta de visualización. En este caso, la columna “name” (marca de auto), será la que represente cada poli-línea del gráfico, y se creará un eje por cada columna restante. Notar que todos los datos son numéricos, a excepción de la primera columna.

Instalación

Para instalar DPC simplemente descomprimir el archivo *dpc-1.1.zip* adjunto en el directorio público de un servidor HTTP tal como Apache, Windows Server, etc. También es posible

utilizar la herramienta en forma local (en un directorio del sistema de archivos local), a excepción de los ejemplos predefinidos que no funcionarán si no se ejecutan mediante un *request* HTTP.

```

1 name,economy (mpg),cylinders,displacement (cc),power (hp),weight (lb),0-60 mph (s),year
2 AMC Ambassador Brougham,13,8,360,175,3821,11,73
3 Audi 100 LS,20,4,114,91,2582,14,73
4 BMW 2002,26,4,121,113,2234,12.5,70
5 Buick Century 350,13,8,350,175,4100,13,73
6 Cadillac Eldorado,23,8,350,125,3900,17.4,79
7 Chevrolet Chevelle Malibu,16,6,250,105,3897,18.5,75
8 Chevrolet Woody,24.5,4,98,60,2164,22.1,76
9 Chevy C10,13,8,350,145,4055,12,76
10 Chrysler Cordoba,15.5,8,400,190,4325,12.2,77
11 Citroen DS-21 Pallas,,4,133,115,3090,17.5,70
12 Datsun PL510,27,4,97,88,2130,14.5,71
13 Dodge Aries SE,29,4,135,84,2525,16,82
14 Fiat 124 Sport Coupe,26,4,98,90,2265,15.5,73
15 Ford Capri II,25,4,140,92,2572,14.9,76
16 Hi 1200D,9,8,304,193,4732,18.5,70
17 Honda Accord CVCC,31.5,4,98,68,2045,18.5,77
18 Maxda GLC Deluxe,34.1,4,86,65,1975,15.2,79
19 Mazda 626,31.3,4,120,75,2542,17.5,80
20 Mercedes-Benz 240D,30,4,146,67,3250,21.8,80
21 Mercury Capri 2000,23,4,122,86,2220,14,71
22 Nissan Stanza XE,36,4,120,88,2160,14.5,82
23 Oldsmobile Cutlass Ciera (Diesel),38,6,262,85,3015,17,82
24 Opel 1900,25,4,116,81,2220,16.9,76
25 Peugeot 304,30,4,79,70,2074,19.5,71
26 Plymouth Arrow GS,25.5,4,122,96,2300,15.5,77
27 Pontiac Astro,23,4,140,78,2592,18.5,75
28 Renault 12 (Wagon),26,4,96,69,2189,18,72
29 Saab 900S,,4,121,110,2800,15.4,81
30 Subaru DL,30,4,97,67,1985,16.4,77
31 Toyota Carina,20,4,97,88,2279,19,73
32 Triumph TR7 Coupe,35,4,122,88,2500,15.1,80
33 Volkswagen Rabbit,29.8,4,89,62,1845,15.3,80
34 Volkswagen 1131 Deluxe Sedan,26,4,97,46,1835,20.5,70
35 Volvo 144EA,19,4,121,112,2868,15.5,73

```

5: Ejemplo de archivo en formato CSV compatible con DPC

Trabajos a futuro

Son muy numerosas las posibilidades de ampliación de la herramienta. Muchas de estas ampliaciones propuestas ya están contempladas y documentadas en el código, especificando el lugar donde deberían implementarse (las mismas están marcadas con la etiqueta “TODO:” en los scripts *Javascript* y código *HTML*). A continuación, se enumera las ampliaciones, que a mi criterio serían las más relevantes, en una potencial siguiente versión de la herramienta.

Soporte para datos ordinales

Ingresar datos, donde una o más dimensiones sean de tipo ordinal (ejemplo: String). D3 ya cuenta con soporte para generar escalas con datos ordinales. Aquí las opciones para implementar esta funcionalidad podrían ser:

1. Pre-procesar los datos de entrada convirtiendo los datos ordinales a nominales. Por ejemplo, asignar un valor numérico a cada *String*. Luego de este pre procesamiento de los datos se podría usar DPC sin ninguna modificación adicional.
2. Cargar los datos en 2 pasos: en el primer paso se ingresan los datos como se implemento actualmente; y en un segundo paso se le permite al usuario, luego de analizar el *input* inicial, editar el tipo de datos que tendrá cada coordenada (ya sea mediante el ingreso de un

pseudo-Javascript o generando inputs HTML dinámicos que permitan al usuario configurar cada dimensión del gráfico a generar). Esto haría a la herramienta sensiblemente más compleja de utilizar.

3. Autodetectar el tipo de cada eje y generar la escala correspondiente, sin agregar ningún paso manual extra.

Soporte para nuevos formatos de entrada

Sería deseable que además de archivos en formato CSV, sea posible ingresar datos en otros formatos conocidos y de uso popular como JSON y XML. D3 ya brinda el soporte para procesar este tipo de archivos con lo cual la inclusión de esta funcionalidad no debería ser muy compleja.

Mejorar el manejo de errores

La herramienta en su versión actual, solo realiza algunas verificaciones básicas sobre los datos de entrada y durante el proceso de generación de la visualización. Sería deseable mejorar la información y análisis, brindado por la herramienta; para el manejo, procesamiento e información de los errores.

Exportar la visualización en diferentes formatos

Exportar el gráfico generado a otros formatos tales como PDF, PNG y SVG. Una posible solución para implementar esta funcionalidad se puede consultar en [13]. El código *HTML* de la herramienta ya tiene incorporada una *toolbar* (oculta en la versión actual) sobre el gráfico para implementar la exportación; solo restaría completar la llamada a cada función que implemente dicha funcionalidad.

Mejorar el styling general de la visualización

Contar con más opciones de visualización para el gráfico de coordenadas paralelas, como por ejemplo:

1. Utilizar un color de línea diferente para cada grupo de datos. Se puede consultar un ejemplo en [14] y [15].
2. Mostrar los datos ingresados en una tabla que permita interactuar con los datos en conjunto con las interacciones ya provistas por el gráfico de coordenadas paralelas. Se puede consultar un ejemplo en [15].
3. Cambiar el color de fondo del gráfico o la página completa, como se puede observar en [15]. Para esta funcionalidad ya están los botones en la *toolbar* oculta en el HTML de la herramienta.

Conclusiones

Las coordenadas paralelas son una técnica muy versátil y útil para encontrar estructuras en conjuntos de datos de tamaño moderado. Con solo un poco de experiencia, es posible reconocer patrones rápidamente e incluso estimar el grado de correlación de los datos, etc.

DPC es una simple herramienta que permite rápidamente visualizar datos con un gráfico de coordenadas paralelas, para entre otras cosas determinar si es una visualización adecuada para los datos con los que se cuenta; sin la penalidad de implementar una visualización completa desde cero. Esta herramienta es un pequeño aporte a la comunidad académica con un gran potencial de ampliación.

Referencias

- [1] eagereyes.org, «Parallel Coordinates,» [En línea]. Available: <https://eagereyes.org/techniques/parallel-coordinates>. [Último acceso: 26 Abril 2015].
- [2] Vyglab, «Vyglab - Visualización y Computación Gráfica,» [En línea]. Available: <http://vyglab.cs.uns.edu.ar/index.php/projects/student-proyectos/call-for/140-propuestas-para-trabajos-finales-de-grado-2>. [Último acceso: 26 Abril 2015].
- [3] <http://d3js.org/>, «D3 Data-Driven Documents,» [En línea]. Available: <http://d3js.org/>. [Último acceso: 26 Abril 2015].
- [4] «Cross-Origin XMLHttpRequest,» [En línea]. Available: <https://developer.chrome.com/extensions/xhr>. [Último acceso: 28 Abril 2015].
- [5] «Bootstrap,» [En línea]. Available: <http://getbootstrap.com/>. [Último acceso: 28 Abril 2015].
- [6] Krajee, «Bootstrap File Input,» [En línea]. Available: <http://plugins.krajee.com/file-input>. [Último acceso: 28 Abril 2015].
- [7] Python Software Foundation, «Python SimpleHTTPServer,» [En línea]. Available: <https://docs.python.org/2/library/simplehttpserver.html>. [Último acceso: 28 Abril 2015].
- [8] jasondavies's blocks, «Parallel Coordinates,» [En línea]. Available: <http://bl.ocks.org/jasondavies/1341281>. [Último acceso: 28 Abril 2015].
- [9] W3.org, «File API - W3C Working Draft 21 April 2015,» [En línea]. Available: <http://www.w3.org/TR/FileAPI/>. [Último acceso: 28 Abril 2015].
- [10] M. Maclean, «D3 Tips and Tricks,» [En línea]. Available: <https://leanpub.com/D3-Tips-and-Tricks/read>. [Último acceso: 28 Abril 2015].
- [11] mbostock's blocks, «Cost of Living - Parallel Coordinates,» [En línea]. Available: <http://bl.ocks.org/mbostock/1558011>. [Último acceso: 28 Abril 2015].
- [12] «RFC4180 - Common Format and MIME Type for Comma-Separated Values (CSV) Files,» [En línea]. Available: <http://tools.ietf.org/html/rfc4180>. [Último acceso: 28 Abril 2015].
- [13] «d3js/SVG Export demo,» [En línea]. Available: <http://d3export.housegordon.org/>. [Último

acceso: 28 Abril 2015].

- [14] C. Evans, «Parallel Coordinates,» [En línea]. Available:
http://www.columbia.edu/~cme2126/datavisuals/bigdata_parallelcoordinates.html. [Último
acceso: 28 Abril 2015].
- [15] «Nutrient Contents - Parallel Coordinates,» [En línea]. Available:
<http://exposedata.com/parallel/>. [Último acceso: 28 Abril 2015].
- [16] S. Few, «Solutions to the Problem of Over-Plotting in Graphs,» [En línea]. Available:
http://www.perceptualedge.com/articles/visual_business_intelligence/over-plotting_in_graphs.pdf. [Último acceso: 28 Abril 2015].